BODYBOUND: A BENCHMARK FOR DIAGNOSING ROBOT-POLICY RESPONSES TO BODY CHANGES BEYOND TASK SUCCESS

Anonymous authors
Paper under double-blind review

ABSTRACT

Task success does not by itself show whether a robot policy adapts when its body changes. A policy may fail before the intervention, or the same action sequence may work on both bodies. BodyBound is a controlled benchmark for separating these cases. It pairs executions with the same scene, goal, initial state, and rollout budget while changing one body property: joint limits, jaw capacity, finger-force limit, or wrist-guard geometry within a Panda topology. An independent trace checker is combined with four diagnostics: familiar-start competence, expert attainability witnesses, reciprocal body-description interventions, and fixed-action replay. We release 100 simulated RGB-D demonstrations and evaluate HPT and Diffusion Policy on 60 fresh paired scenes (720 matching-description rollouts). On the public test split, success is 4.2% for HPT and 19.4% for Diffusion Policy; all six checkpoints obtain zero successes in Scene OOD. The learned policies also fail on many familiar starts, limiting attribution of failures to body changes. In a competent Portal reference, all 16 fresh pairs succeed with zero description benefit and shared successful actions; a Grasp reference shows a positive description benefit despite shared solutions. BodyBound therefore measures competence, physical response, and evidence of adaptation separately.

1 INTRODUCTION

A robot may need a different grasp when its gripper opening changes, or a different path when a wrist guard becomes wider. A successful rollout does not reveal whether the policy made that adjustment: the same commands may work on both bodies. A failure is ambiguous for the opposite reason. A policy that already fails on a familiar task may fail again after the body changes, even when the changed body can complete it. Adaptation therefore requires an evaluation that links outcomes to controlled physical interventions and to the actions that produced them.

Manipulation benchmarks measure task and environment transfer, and morphology benchmarks vary the robot itself (Liu et al., 2023; Nasiriany et al., 2026; Parakh et al., 2025). They do not, by task success alone, answer three different questions: can the policy solve the task, does the physical change alter its behavior, and is there evidence that it adapted to that change? BodyBound makes the first question explicit and uses paired scenes to study the other two. Its basic unit holds the scene, goal, initial state, rollout seed, checkpoint, and command budget fixed while one body property changes within the Panda topology.

The benchmark combines a visual demonstration dataset with controls that target these questions (Figure 1). Familiar starts measure competence under training-supported conditions, and successful expert executions provide attainability witnesses. Reciprocal body descriptions test whether the supplied description changes outcomes, while replaying one command sequence on both bodies tests whether the task actually requires different successful actions. A policy can benefit from a matching description even when a shared solution exists.

We release 100 RGB-D demonstrations and evaluate HPT and Diffusion Policy on 60 fresh paired scenes. Public-test success is 4.2% and 19.4%, respectively; both methods fail throughout Scene OOD, and familiar-start results show limited competence before a body shift. A public-input Portal reference succeeds on all 16 fresh pairs with no matching-description benefit and shared successful

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

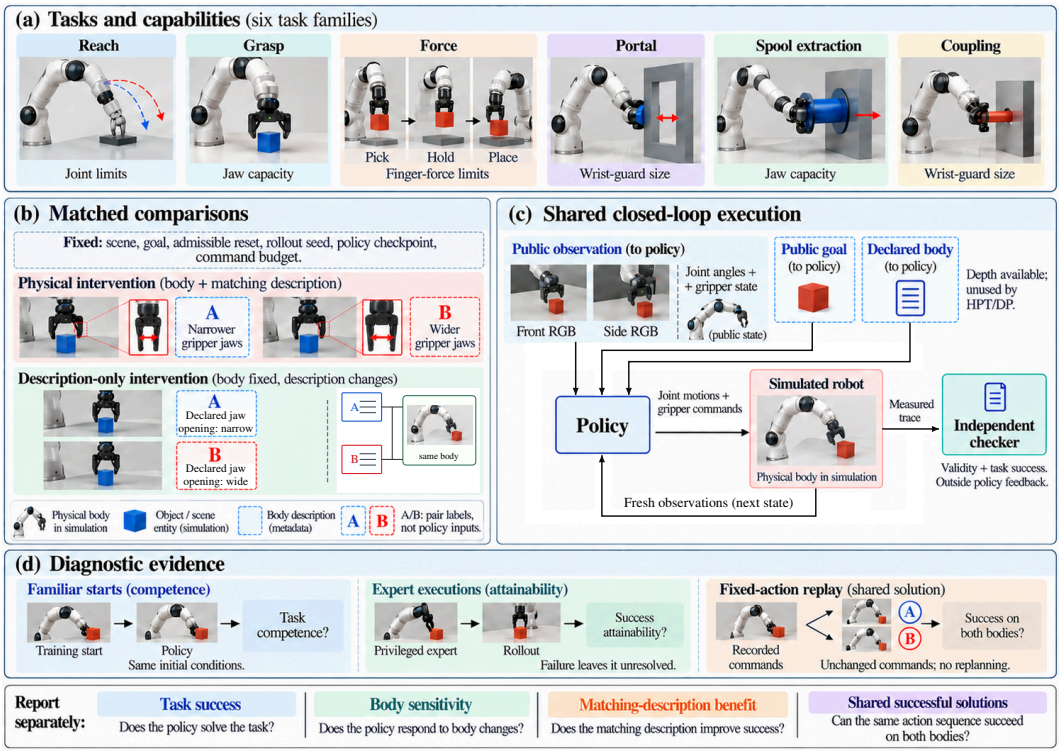

Figure 1: BodyBound in simulation. **(a)** Task schematics show six families and their intervened body parameters. **(b)** Paired comparisons change the physical body with its matching description, or the description alone. **(c)** Policies act from public observations, goals, and body descriptions; fresh observations close the loop, while an independent checker evaluates measured traces. HPT and Diffusion Policy use RGB, not depth. **(d)** Familiar starts, expert executions, and unchanged-command replay separate task competence, attainability, and shared solutions. Scene geometry stays fixed within each pair.

commands. A Grasp reference supplies the complementary case: a matching description helps even when shared commands succeed.

We introduce paired body interventions with independently checked execution, release a synchronized visual dataset with explicit task–body coverage, and show how competence, physical response, and action necessity provide separate evidence.

2 Related Work

Manipulation benchmarks and data. Manipulation benchmarks evaluate generalization across tasks, goals, and scenes through multi-task and language-conditioned control (Yu et al., 2020; James et al., 2020; Mees et al., 2022; Liu et al., 2023). Simulation frameworks broaden the tasks and environments available for training and evaluation (Tao et al., 2025; Nasiriany et al., 2026), while Open X-Embodiment pools demonstrations across robot platforms (O'Neill et al., 2024). These resources support learning from varied experience and measuring transfer to new settings. BodyBound complements this coverage with comparisons within a fixed scene: the goal, initial joint state, and rollout budget remain constant while one generating body parameter changes within the Panda topology. This pairing makes the physical intervention explicit when interpreting differences in policy behavior and success.

Control across bodies. Hardware-conditioned, graph-based, modular, and transformer policies encode physical information or robot structure to share control across embodiments (Chen et al., 2018; Wang et al., 2018; Huang et al., 2020; Gupta et al., 2022). Cross-embodiment benchmarks

and transfer methods examine how learned behavior carries across robots (Parakh et al., 2025; Sridhar et al., 2026). Two close connections concern the controls used to interpret this transfer. URMA perturbs joint descriptions to test policy conditioning (Bohlinger et al., 2025), and SHADOW replays actions across embodiments with matched initial object positions (Lepert et al., 2025). BodyBound combines physical interventions, reciprocal descriptions, and fixed-command replay on paired bodies. These comparisons distinguish sensitivity to the supplied description from the existence of shared successful commands. Both inform adaptation, but a description benefit can coexist with a sequence that succeeds on both bodies.

Policies and evaluation. Action chunking and diffusion modeling provide approaches to visual imitation (Zhao et al., 2023; Chi et al., 2023); HPT and CrossFormer study learning from heterogeneous robot data (Wang et al., 2024; Doshi et al., 2025). We evaluate HPT and Diffusion Policy with matched observations, demonstrations, and rollout budgets. Controlled evaluation offers a complementary perspective: CausalWorld, RoboEval, and LIBERO-Plus examine perturbations or capability dimensions beyond aggregate task success (Ahmed et al., 2021; Wang et al., 2026; Fei et al., 2026). BodyBound applies this perspective to failure attribution under body changes. Familiar-start tests assess competence on demonstrated initial conditions, and successful expert executions provide attainability witnesses. Description interventions and action replay further place outcomes in the context of policy conditioning and shared solutions across bodies.

3 THE BODYBOUND DATASET AND BENCHMARK

3.1 PAIRED TASKS AND PHYSICAL INTERVENTIONS

Each panel contains one scene and two bodies. The six task families cover reaching, grasping, force-limited placement, portal crossing, portal–grasp coupling, and spool extraction. The checker evaluates contacts, motion, placement, and passage constraints from the measured rollout; full predicates are in Appendix A.

Each pair changes one generating body parameter. Joint and jaw limits change available motion, finger-force limits change the native actuator cap, and guard width changes visual and collision geometry together with mass and inertia at fixed density. All variants retain the Panda topology; exact ranges are listed in Appendix A.

Observations, actions, and outcomes. The public interface provides front and side RGB-D, proprioception, the public goal, and a numerical body description. Private object poses, rewards, body IDs, and checker outcomes are withheld from the policy. Both learned baselines use the two RGB views without depth and issue absolute targets for seven joints and total jaw opening at 20 Hz, with a 400-command episode budget. The remaining sensor and simulator settings are in Appendix A.1.

Task success is determined from measured contacts and motion, not commanded targets. An independent checker recomputes each predicate from the saved trace outside the policy feedback loop; execution validity and task success are recorded separately. Reset geometry is checked independently of outcome agreement.

3.2 DEMONSTRATIONS, COVERAGE, AND QUALITY

The learning dataset contains 100 simulated expert demonstrations: 72 training episodes from 47 panels and 28 demonstration-development episodes from 18 panels. Each episode contains synchronized observations, robot state, body and goal descriptions, actions, and skill annotations; the expert's private planning state is excluded from policy inputs. Coverage is uneven, and Spool has no demonstrations. Table 1 summarizes the split; action-window counts and trajectory overlap are given in Appendix B.

We group paired bodies before splitting, reject initial collisions, and recompute outcomes from saved traces. A reset-overlap issue in Reach was corrected before training; all 32 corrected demonstrations pass synchronized-data, checker, and replay checks, and the fixed expert succeeds on 20/20 corrected starts versus 0/20 for a fixed-target control. All 120 fresh members pass reset admission. Appendix B.1 gives the correction record; Figure 2 shows native observations.

Table 1: Demonstration coverage. Body counts denote distinct physical specifications within a family; the total is their union. Every evaluation condition includes four members per family, including Spool.

Family	Train demos	Dev demos	Train bodies	Eval. members/split
Reach	24	8	2	4
Grasp	10	5	1	4
Force	12	4	1	4
Portal	14	7	2	4
Spool extraction	0	0	0	4
Grasp–portal coupling	12	4	2	4
Total	72	28	6	24

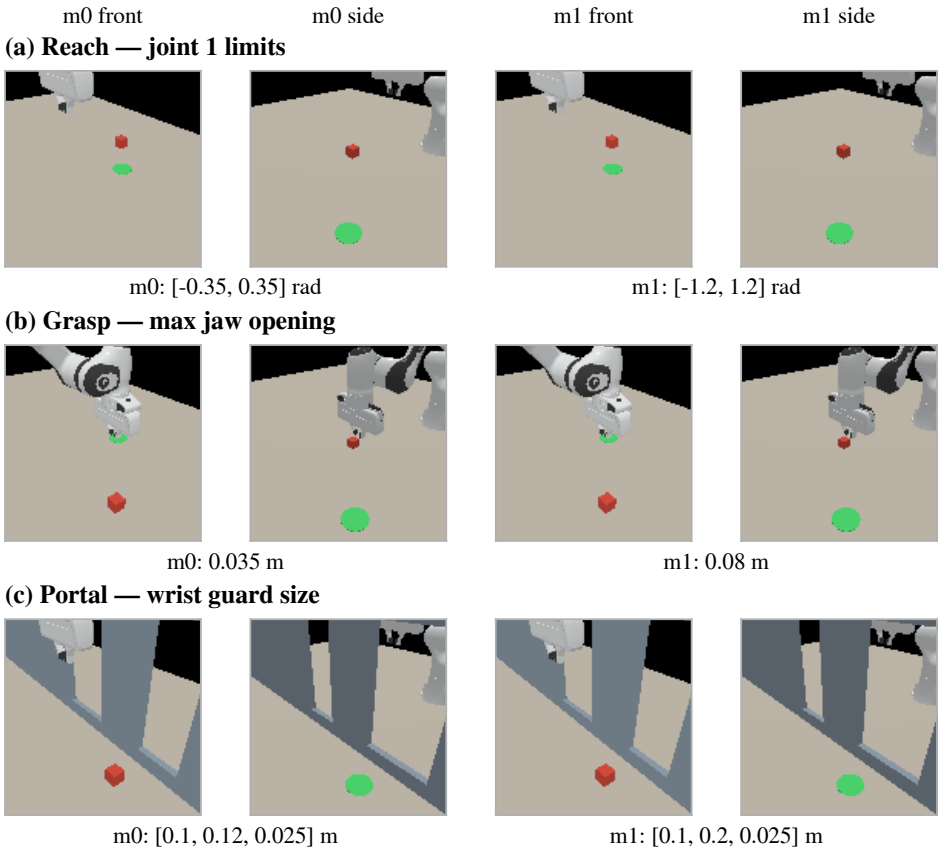

Figure 2: Native initial RGB observations from the released interface. Each body has front and side views at 128×128 resolution; initial images can look alike even when execution capabilities differ.

3.3 Evaluation cohorts

The challenge suite contains 60 fresh paired scenes and 120 physical members. Five conditions each have two scenes per family: Fresh Dev, public test, Body OOD, Scene OOD, and Body+Scene OOD. All five are evaluation-only; demonstration-development loss selects checkpoints. Scene shifts move object and goal offsets beyond the nominal lateral range. Body OOD shifts the registered parameter pairs, though some shifted bodies occur in training. Different conditions use different scenes.

The 60 pairs are sampled before policy execution, and every selected member remains in the denominator, including members without demonstrations or successful expert attempts. The challenge suite is analyzed separately from the Portal reference with competent public-input controllers.

Table 2: Fresh-scene success (%, mean ± sample SD over three seeds). Each seed evaluates the same 24 members per condition. SD describes seed variation, not a confidence interval. Fresh Dev is evaluation-only; Body OOD is a nominal-parameter shift, not uniformly unseen training bodies.

Method	Fresh Dev	Public test	Body OOD	Scene OOD	Body+Scene OOD
HPT	2.8 ± 4.8	4.2 ± 4.2	1.4 ± 2.4	0.0 ± 0.0	1.4 ± 2.4
Diffusion Policy	20.8 ± 4.2	19.4 ± 6.4	16.7 ± 4.2	0.0 ± 0.0	2.8 ± 2.4

4 Policies and evaluation measures

HPT uses a frozen pretrained transformer and image encoder with trainable domain stems and a diffusion action head (Wang et al., 2024). Diffusion Policy trains its visual action model from the demonstrations with ImageNet-initialized encoders (Chi et al., 2023). Both methods receive the same RGB, proprioception, goal, and body inputs and predict absolute joint/jaw actions. We train three seeds (13, 37, and 71) for 50 epochs; demonstration-development loss selects the checkpoint. Their complete recipes are in Appendix C.

Success and paired response. Let $y_{s,p,b}$ denote checked success for training seed s, scene p, and body $b \in \{0, 1\}$ under its matching description. For the 12 panels $\mathcal{P}$ in a condition and the two panels $\mathcal{P}_f$ in family f, we report

$$R_s = \frac{1}{2|\mathcal{P}|} \sum_{p \in \mathcal{P}} \sum_{b=0}^{1} y_{s,p,b}, \qquad \Delta_f = \frac{1}{3|\mathcal{P}_f|} \sum_{s=1}^{3} \sum_{p \in \mathcal{P}_f} (y_{s,p,1} - y_{s,p,0}). \tag{1}$$

The mean and sample SD of R_s describe variation across checkpoints on fixed scenes. $100\Delta_f$ is the signed within-scene response in percentage points. Member 1 increases the designated parameter, which may enlarge available motion or a collision envelope; the physical body and its matching description change together.

Matching-description benefit and shared solutions. For a fixed checkpoint and scene, let Y_{bd} be success on physical body b with supplied description d. The complete 2×2 crossing gives

$$C = \frac{Y_{00} + Y_{11} - Y_{01} - Y_{10}}{2}, \qquad P_d = Y_{1d} - Y_{0d}. \tag{2}$$

C measures the average benefit of the matching description; P_d measures the physical response at a fixed description while allowing feedback. A zero C can result from common failure, common success, or cancellation, so we inspect the four cells. Fixed-action replay asks a separate question: an unchanged sequence that succeeds on both bodies is evidence that different successful actions are unnecessary for that reset.

The paired scene is the sampling unit. Three seeds repeat the same scenes; they do not triple scene diversity. We use descriptive seed variation and scene-resampling intervals for the factorial results, with the complete procedure in Appendix E.1.

5 Results

5.1 Task success and body response

The six checkpoints complete 720 matching-description rollouts: 24 members per condition across five conditions. Table 2 summarizes the full cohort; the appendix gives seed- and family-level results.

HPT averages 4.2% public-test success (0–2 successes per seed), while Diffusion Policy averages 19.4% (3–6 per seed). Both fail throughout Scene OOD. Success is concentrated in Reach and Grasp: Diffusion Policy reaches 83.3% and 33.3% on those public-test families, whereas HPT succeeds only on those two families; neither method succeeds on Spool. Body OOD is not uniformly

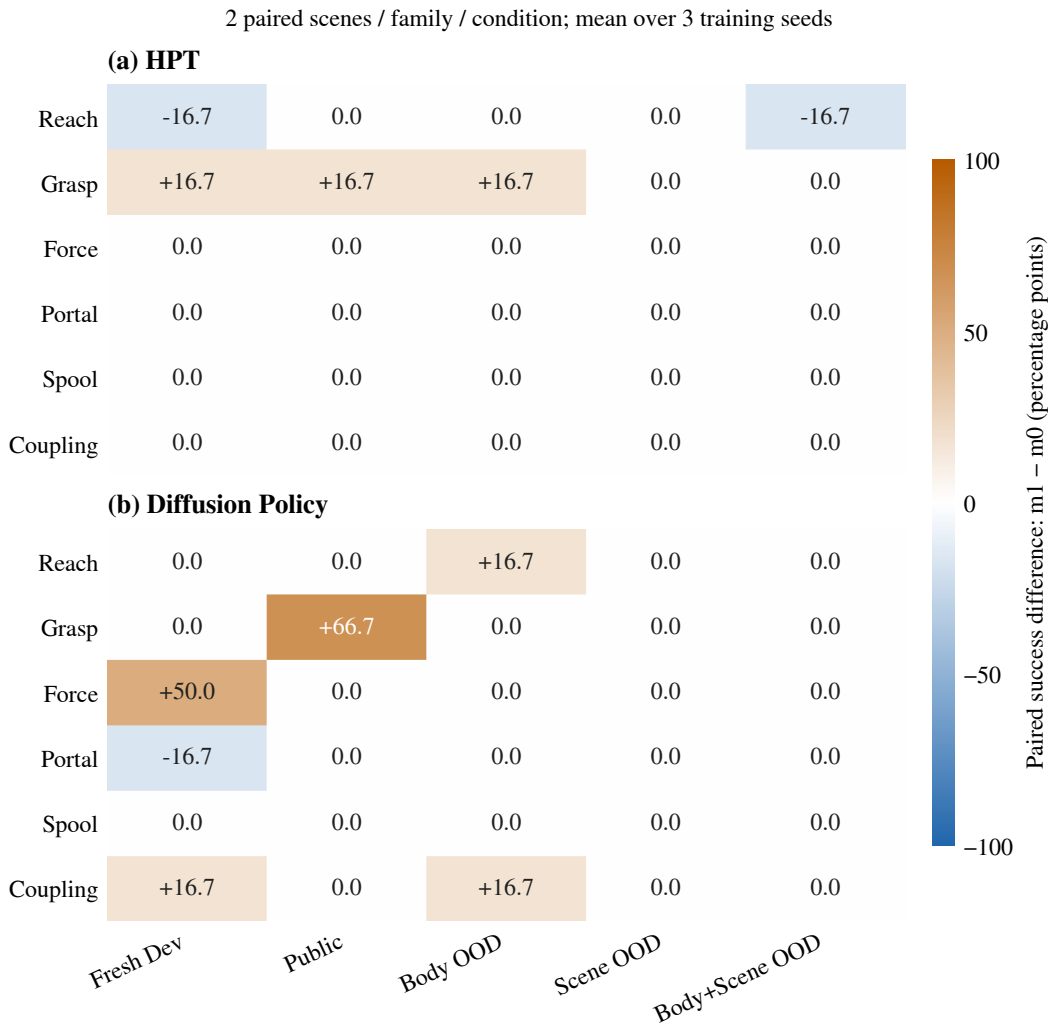

Figure 3: Within-scene paired response, $100\Delta_f$, in percentage points. Each cell averages two shared scenes and three checkpoints. Physical body and matching description change together; these are not fixed-description effects P_d. Member 1 increases the designated parameter. The common scale retains all 60 values, including zeros caused by common failure or cancelling seed-level responses.

harder than the other conditions, and its comparison with Scene OOD is descriptive because the scenes differ.

Figure 3 compares the two bodies within each scene. Eight of the 60 method–family–condition means are positive, three are negative, and 49 are zero. Diffusion Policy reaches $+66.7$ points on public-test Grasp and -16.7 on Fresh Dev Portal. Equal aggregate success can therefore hide different body responses; the complete seed-level values are in Appendix D.

5.2 Are failures specific to changing the body?

Familiar-start competence. We select eight successful training demonstrations before policy execution, one per supported task-family/exact-body stratum. All six checkpoints start from matching histories, goals, and body descriptions. HPT succeeds on 1/8, 1/8, and 0/8 starts; Diffusion Policy on 4/8, 3/8, and 2/8. The failures occur under familiar conditions, before any new-body transfer is involved.

Attainability and coverage. The fixed privileged expert succeeds on 57/120 fresh members, including all 20 Reach members and no Spool member. On witnessed members, HPT achieves 7/171 successes and Diffusion Policy 40/171 across three seeds. Diffusion Policy also succeeds on two expert-failed members, so expert failure does not define task impossibility. The expert uses privileged geometry and serves as a witness source, not an equal-information baseline.

5.3 Does the body description help, and must actions differ?

The 720 matching and 720 reciprocal-description rollouts fill all 1,440 factorial cells. HPT's mean matching benefit is -0.6 percentage points (empirical 95% resampling interval $[-1.1, 0.0]$); Diffusion Policy's is -0.8 ($[-1.9, 0.3]$). Diffusion Policy nevertheless has physical contrasts $P_0 = +6.1$ and $P_1 = +4.4$ points. Physical response and description benefit therefore need not agree. Because both policies still observe RGB and proprioception, a near-zero description effect is not evidence that they ignore the body.

Recorded-command replay finds shared solutions on the same challenge scenes. Of 110 expert sequences compatible with both bodies' ranges, 36 succeed unchanged on both bodies, covering 18 panels. Ten other sequences fail after body-range projection. These shared successes show that a body change need not require different successful commands for a given reset.

The complete panel analysis, including positive, negative, and zero blocks, appears in Appendix D.2.

5.4 Calibration with competent public-input references

The analytic controllers are interpretation references, not competitive baselines. Controller A uses the supplied body description, while S uses a constant description. On the Portal reference, all 128 body–description cells succeed on 16 fresh pairs; both controllers have $C = 0$, and all 16 fixed S_{00} sequences replay successfully on both bodies. This is the competent case: neither a matching-description benefit nor different commands is needed.

The Grasp reference gives the complementary case. Across 12 scenes, A has a 50.0-point description effect and S has 0.0 points. Ten scenes have shared successful commands even when A benefits from the matching description. Only 22/24 physical members have success witnesses; the full protocols and results are in Appendix F.

6 Limitations

BodyBound uses one simulated Panda topology; hardware and cross-platform transfer remain untested. The 100 demonstrations are imbalanced and exclude Spool. With two scenes per family and condition, factorial and support analyses remain descriptive. Familiar-start competence is limited. Portal and Grasp references calibrate interpretation without extending learned-policy evidence. Fixed training budgets leave the full potential of HPT and Diffusion Policy unmeasured.

7 Conclusion

BodyBound turns body changes into paired interventions that can be examined through measured execution. Its results show why success must be read together with competence, physical response, and action replay: learned policies already fail on familiar starts, physical responses can vary even when success is similar, and a matching description can help without making different actions necessary. These distinctions are the benchmark's central contribution.

Reproducibility statement

The anonymous supplementary material provides the corrected demonstrations, task and checker specifications, training and evaluation procedures, and execution records. The selected checkpoints and any non-redistributable pretrained components are identified there together with their source information. Appendix C gives the training recipes and Appendix G describes the reproducibility checks.

AI USE STATEMENT

No generative AI tools were used in this work.

REFERENCES

Ossama Ahmed, Frederik Träuble, Anirudh Goyal, Alexander Neitz, Manuel Wüthrich, Yoshua Bengio, Bernhard Schölkopf, and Stefan Bauer. CausalWorld: A robotic manipulation benchmark for causal structure and transfer learning. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.04296.

Nico Bohlinger, Grzegorz Czechmanowski, Maciej Piotr Krupka, Piotr Kicki, Krzysztof Walas, Jan Peters, and Davide Tateo. One policy to run them all: an end-to-end learning approach to multi-embodiment locomotion. In *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 3356–3378. PMLR, 2025. URL https://proceedings.mlr.press/v270/bohlinger25a.html.

Tao Chen, Adithyavairavan Murali, and Abhinav Gupta. Hardware conditioned policies for multi-robot transfer learning. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/b8cfbf77a3d250a4523ba67a65a7d031-Abstract.html.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin CM Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Proceedings of Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.026. URL https://www.roboticsproceedings.org/rss19/p026.html.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014. doi: 10.3115/v1/D14-1179. URL https://aclanthology.org/D14-1179/.

Ria Doshi, Homer Rich Walke, Oier Mees, Sudeep Dasari, and Sergey Levine. Scaling cross-embodied learning: One policy for manipulation, navigation, locomotion and aviation. In *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 496–512. PMLR, 2025. URL https://proceedings.mlr.press/v270/doshi25a.html.

Senyu Fei, Siyin Wang, Junhao Shi, et al. LIBERO-Plus: A progressive robustness benchmark for visual-language-action models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 38574–38583, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Fei_LIBERO-Plus_A_Progressive_Robustness_Benchmark_for_Visual-Language-Action_Models_CVPR_2026_paper.html.

Agrim Gupta, Linxi Fan, Surya Ganguli, and Li Fei-Fei. MetaMorph: Learning universal controllers with transformers. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=Opmqtk_GvYL.

Wenlong Huang, Igor Mordatch, and Deepak Pathak. One policy to control them all: Shared modular policies for agent-agnostic control. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4455–4464. PMLR, 2020. URL https://proceedings.mlr.press/v119/huang20d.html.

Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J. Davison. RLBench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters*, 5(2):3019–3026, 2020. doi: 10.1109/LRA.2020.2974707. URL https://doi.org/10.1109/LRA.2020.2974707.

James J. Kuffner and Steven M. LaValle. RRT-Connect: An efficient approach to single-query path planning. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 2, pp. 995–1001, 2000. doi: 10.1109/ROBOT.2000.844730. URL https://lavalle.pl/papers/KufLav00.pdf.

Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. RMA: Rapid motor adaptation for legged robots. In *Proceedings of Robotics: Science and Systems*, 2021. doi: 10.15607/RSS.2021.XVII.011. URL https://www.roboticsproceedings.org/rss17/p011.html.

Marion Lepert, Ria Doshi, and Jeannette Bohg. SHADOW: Leveraging segmentation masks for cross-embodiment policy transfer. In *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 3536–3550. PMLR, 2025. URL https://proceedings.mlr.press/v270/lepert25a.html.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. LIBERO: Benchmarking knowledge transfer for lifelong robot learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 44776–44791, 2023. doi: 10.52202/075280-1939. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/8c3c666820ea055a77726d66fc7d447f-Abstract-Datasets_and_Benchmarks.html.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Oier Mees, Lukas Hermann, Erick Rosete-Beas, and Wolfram Burgard. CALVIN: A benchmark for language-conditioned policy learning for long-horizon robot manipulation tasks. *IEEE Robotics and Automation Letters*, 7(3):7327–7334, 2022. doi: 10.1109/LRA.2022.3180108. URL https://doi.org/10.1109/LRA.2022.3180108.

Soroush Nasiriany, Sepehr Nasiriany, Abhiram Maddukuri, and Yuke Zhu. RoboCasa365: A large-scale simulation framework for training and benchmarking generalist robots. In *International Conference on Learning Representations*, pp. 98643–98667, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/a05003fdb1e9562ab0c0a9719ea4de10-Abstract-Conference.html.

Abby O'Neill, Abdul Rehman, Abhiram Maddukuri, et al. Open X-embodiment: Robotic learning datasets and RT-X models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6892–6903. IEEE, 2024. doi: 10.1109/ICRA57147.2024.10611477.

Meenal Parakh, Alexandre Kirchmeyer, Beining Han, and Jia Deng. AnyBody: A benchmark suite for cross-embodiment manipulation. *arXiv preprint arXiv:2505.14986*, 2025. doi: 10.48550/arXiv.2505.14986. URL https://arxiv.org/abs/2505.14986.

Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3803–3810. IEEE, 2018. doi: 10.1109/ICRA.2018.8460528.

Mohit Shridhar, Lucas Manuelli, and Dieter Fox. Perceiver-Actor: A multi-task transformer for robotic manipulation. In *Proceedings of The 6th Conference on Robot Learning*, volume 205 of *Proceedings of Machine Learning Research*, pp. 785–799. PMLR, 2023. URL https://proceedings.mlr.press/v205/shridhar23a.html.

Ajay Sridhar, Jensen Gao, Jonathan Yang, Jean Mercat, Suneel Belkhale, and Dorsa Sadigh. Cross-embodiment transfer via behavior-aligned representations. In *International Conference on Robotics and Automation (ICRA)*, 2026. URL https://ajaysridhar.com/barx/.

Stone Tao, Fanbo Xiang, Arth Shukla, et al. Demonstrating GPU parallelized robot simulation and rendering for generalizable embodied AI with ManiSkill3. In *Proceedings of Robotics: Science and Systems*, 2025. doi: 10.15607/RSS.2025.XXI.021. URL https://www.roboticsproceedings.org/rss21/p021.html.

Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 23–30. IEEE, 2017. doi: 10.1109/IROS.2017.8202133.

Lirui Wang, Xinlei Chen, Jialiang Zhao, and Kaiming He. Scaling proprioceptive-visual learning with heterogeneous pre-trained transformers. In *Advances in Neural Information Processing Systems*, volume 37, pp. 124420–124450, 2024. doi: 10.52202/079017-3952. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/e0f393e7980a24fd12fa6f15adfa25fb-Abstract-Conference.html.

Tingwu Wang, Renjie Liao, Jimmy Ba, and Sanja Fidler. NerveNet: Learning structured policy with graph neural networks. In *International Conference on Learning Representations*, 2018. URL https://iclr.cc/virtual/2018/poster/66.

Yi Ru Wang, Carter Ung, Christopher Tan, et al. RoboEval: Where robotic manipulation meets structured and scalable evaluation. *arXiv preprint arXiv:2507.00435v2*, 2026. URL https://arxiv.org/abs/2507.00435v2. Version 2 revised May 4, 2026; first posted July 1, 2025.

Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. Meta-World: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Proceedings of the Conference on Robot Learning*, volume 100 of *Proceedings of Machine Learning Research*, pp. 1094–1100. PMLR, 2020. URL https://proceedings.mlr.press/v100/yu20a.html.

Yanjie Ze, Gu Zhang, Kangning Zhang, Chenyuan Hu, Muhan Wang, and Huazhe Xu. 3D diffusion policy: Generalizable visuomotor policy learning via simple 3D representations. In *Proceedings of Robotics: Science and Systems*, 2024. doi: 10.15607/RSS.2024.XX.067. URL https://www.roboticsproceedings.org/rss20/p067.html.

Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. In *Proceedings of Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.016. URL https://www.roboticsproceedings.org/rss19/p016.html.

A TASK DEFINITIONS AND EXECUTION INTERFACE

The following specifications define the released tasks and the quantities used by the outcome checker. A qualifying success event is remembered for the rest of the rollout. Geometric comparisons use a numerical tolerance of 10^{-9}; contact force is actor-pair net impulse divided by the physics timestep. Spool extraction has no prescribed jaw-opening sequence.

Table 3: Task-success predicates. The lower rows define the grasp, hold, placement, and crossing conditions used by the six families.

Family / predicate	Implemented requirement
Reach	At least one finger contacts the fixed pad with force magnitude ≥ 0.5 N for 0.5 s consecutively.
Grasp	Hold predicate H for 2 s consecutively; placement is not required.
Force	H achieved for 2 s, followed by placement predicate P for 0.5 s consecutively.
Portal	Accepted crossing C, followed by P for 0.5 s consecutively; no wall force > 0.5 N before first success.
Spool extraction	Accepted reverse crossing C (far to near side), followed by P for 0.5 s consecutively; no wall force > 0.5 N before first success. Registered order is `free`.
Grasp–portal coupling	Accepted crossing C for the elongated bar, followed by P for 0.5 s consecutively; no wall force > 0.5 N before first success.
Grasp G	Both finger–object contact magnitudes ≥ 0.5 N, each within 85 degrees of the finger's signed outward y-axis.
Hold H	G, object COM lifted by ≥ 60 mm from reset, and every non-finger support force ≤ 0.5 N.
Placement P	Object COM within 45 mm of goal in XY; lowest object point within 5 mm of goal height; every robot–object force ≤ 0.5 N; linear speed ≤ 30 mm/s and angular speed ≤ 0.3 rad/s.
Crossing C	Maintain G during a source-to-destination wall-slab transit. Every sampled section of the object, hand, fingers, and guard (if present) must fit one common aperture. All carried shapes end beyond the destination slab face; unsampled side-to-side transfers are rejected. For extraction, shapes already on the destination side at grasp start need not cross again.
Sampling / score	Predicates are sampled at 100 Hz (50 stable-contact/placement samples; 200 hold samples). Actions run at 20 Hz for at most 400 commands. Success remains true after its first qualifying sample.

Table 4: Body parameters. Each pair changes one generating parameter, including its induced physical properties. I and E denote interpolation and extrapolation relative to the nominal pair, not training exposure.

Body parameter	Nominal (m_0, m_1)	Shifted pair (m_0, m_1)	Relation
Joint 1 half-range (rad)	[0.35, 1.2]	[0.2, 0.7]	E–, I
Maximum jaw opening (mm)	[35, 80]	[25, 110]	E–, E+
Per-finger force limit (N)	[0.5, 60]	[0.25, 35]	E–, I
Guard y-extent (mm)	[120, 200]	[150, 240]	I, E+

I: between nominal pair values; E–/E+: below/above their range. Joint limits are symmetric about zero. Guard x/z extents remain 100/25 mm, origin [0, 0, -20] mm, density 2700 kg/m^3. These labels do not assert training exposure.

A.1 SIMULATOR AND PUBLIC INPUTS

At each policy query, observations contain the two most recent control ticks, not the two most recent queries. At reset, the current snapshot fills both history slots with validity mask $[0, 1]$. Table 6 gives the native settings and sensor shapes. The body description contains topology, numeric collision shapes, masses and inertias, joint ranges, jaw travel, guard geometry, and drive settings. Its compact encoding is accepted only after reconstructing and validating the complete public description against the fixed Panda template.

Table 5: Nominal and shifted scene intervals, relative to each family template in the robot base frame. Body OOD retains nominal scene ranges; Body+Scene OOD combines the two shifts.

Scene parameter	Nominal / body shift	Scene / combined shift
Object x-offset (mm)	[-25, 25]	[-25, 25]
Object y-offset (mm)	[-30, 30]	[80, 120]
Goal x-offset (mm)	[-25, 25]	[-25, 25]
Goal y-offset (mm)	[-30, 30]	[80, 120]
Object scale, all families	[0.9, 1.1]	[0.9, 1.1]
Object mass (g), all except force	[120, 180]	[120, 180]
Object mass (g), force	[320, 480]	[320, 480]

Offsets are relative to each family's object/goal template. The y shift is separated above the nominal interval; x, scale, and mass ranges are unchanged. Body+Scene OOD combines the registered body and scene shifts.

Table 6: Shared simulation and public interface. Shapes omit the batch dimension. Actuator limits and measured contact forces are different quantities.

Component	Setting
Simulation	ManiSkill with SAPIEN/PhysX, CPU dynamics, and CUDA rendering; 100 Hz physics, 20 Hz commands.
Contact solver	Contact/rest offsets 0.002/0 m; 20 position and 4 velocity iterations; TGS and persistent contact manifolds; CCD disabled. Scene-actor static/dynamic friction 1.0/1.0, restitution 0. Native finger friction coefficients were not separately recorded.
Drives	Arm stiffness/damping 1000/100 in N m/rad and N m s/rad, per-joint cap 100 N m. Finger stiffness/damping 1000/100 in N/m and N s/m; body-specific per-finger cap. Robot-link gravity disabled; object gravity 9.81 m/s^2.
Cameras	Front and side RGB: uint8 $[2, 128, 128, 3]$; depth (m) and validity $[2, 128, 128, 1]$; 3×3 intrinsics, 4×4 camera-to-base transforms; vertical FOV 50°, near/far 0.01/10 m.
Robot state	Two 15-vectors: seven joint positions, seven velocities, and total jaw opening.
Public goal	Instruction and visible-target text, limited to 512 and 256 characters. Private target coordinates are withheld.
Actions	Eight absolute targets: seven angles (rad) and total jaw opening (m). Each finger receives half the jaw target. Global command bounds do not rescale with the body.

The ordered 24-dimensional body vector contains seven lower and seven upper joint limits, maximum jaw opening, per-finger force limit, a guard-presence mask, guard size (x, y, z), mount origin (x, y, z), and density. The mask and seven guard values are zero when absent. Public-description reconstruction uses absolute/relative tolerances $10^{-9}/10^{-7}$. HPT encodes canonical UTF-8 goal JSON with a trained byte-token stem; Diffusion Policy uses marked UTF-8 fields, a 64-dimensional byte embedding, and a 128-dimensional GRU. Neither goal adapter supplies pretrained language knowledge. Depth and calibration are recorded but unused by these two networks.

Geometry. The robot base is at $(-0.615, 0, 0)$ m in world coordinates. Reach uses a 3.5 cm fixed cube; Grasp, Force, and Portal use 4 cm cubes. Coupling uses a $16 \times 3.5 \times 3$ cm bar. The spool has a $6 \times 5 \times 5$ cm center and two $1 \times 11.5 \times 11.5$ cm plates at local $x = \pm 3.5$ cm. Portal and Coupling apertures are 16 and 24 cm wide at $y = -20$ and $+20$ cm, spanning $z =$ 6–48 cm. Spool uses a 13.5 cm opening at $y = 0$, spanning $z =$ 9–28 cm. The wall lies at $x = 2$ cm with thickness 2.5 cm.

Scene separation. For horizontal object and goal coordinates in metres, the scene group is

$$g = \big(\lfloor x_o/0.01 \rfloor, \lfloor y_o/0.01 \rfloor, \lfloor x_g/0.01 \rfloor, \lfloor y_g/0.01 \rfloor\big).$$

The integer tuple is independent of family, body, object scale and mass, height, and predicate thresholds. Generation excludes shared groups across the complete split plan. Fresh evaluation also excludes retained historical groups and exact task/body identities after removing the reset joint vector. This is grid-cell exclusion, not a minimum-distance guarantee. Both members share the task and rollout seed and remain in the same split.

Force interpretation. The body parameter caps the native finger actuator. The 0.5 N contact predicate thresholds measured net contact force, including the interaction response. Comparing these two numbers alone cannot determine attainability; success is evaluated from the executed trace.

B Dataset construction and quality controls

The corrected learning release contains 72 training and 28 demonstration-development episodes from 47 and 18 semantic panels. Each stores synchronized observations, calibrated RGB-D, proprioception, public goal and body description, commands, and skill annotations. Training windows use episode boundaries, not skill boundaries, with repeat-last padding at the end. This yields 17,045 training and 6,757 development windows. Counts refer to action starts within the 100 trajectories.

B.1 Reset correction and demonstration validation

An earlier Reach reset placed robot collision geometry inside the fixed target. A step-zero box-intersection check identified overlap in two of 24 training and two of eight development Reach demonstrations. Replaying commands and obtaining checker agreement did not detect this initialization defect. We therefore corrected all Reach starts uniformly rather than selecting episodes by success.

The corrected joint vector is $(0, -0.35, 0, -2.3, 0, 1.95, \pi/4, 0.01, 0.01)$, with arm angles in radians and finger positions in metres. Scene geometry, goals, bodies, split assignments, seeds, and budgets are unchanged. At the measured native reset, each of 17 registered robot collision primitives must have a separating world-axis projection gap of at least 5 mm from every target part. Outward-padded intervals make this a conservative admission test. Native collision queries also check self-collision under the robot's SRDF exclusions and world geometry including the table-mount cutout.

All 22 reset-only test starts (20 evaluation and two training starts) and all 32 Reach demonstration requests pass admission. The first training panel, with both bodies, forms a pilot of two 102-command trajectories with 103 synchronized observations each. After RGB-D inspection, independent outcome checks, and same-command CPU replay, the remaining 30 requests are collected once each. All 32 first attempts pass the numerical, checker, and replay checks, with no replacement. First, middle, and final RGB frames are inspected for eight episodes: both bodies of the first panel in each generation-seed/split stratum. This is sampled visual inspection, not an exhaustive human annotation of every frame.

The release combines these 32 corrected trajectories with 68 non-Reach episodes whose data and split assignments were unchanged. All reported HPT and DP checkpoints train on this cohort. On the 20 corrected evaluation starts, the original expert completes every task in 101–105 commands; holding the measured initial targets for 400 commands succeeds on none. This control removes the observed fixed-target solution while preserving attainable tasks. Scores from the earlier reset condition are excluded from the reported results.

B.2 Fresh-scene sampling

The 60-scene evaluation is sampled with a fixed seed before policy execution. Proposals are compared with 21,617 historical requests and the 100 corrected demonstration sources under the scene-group rule. Five of 65 proposals are rejected solely for overlap; outcomes do not enter selection. All 120 selected members pass command-free reset admission. These checks validate initial geometry, not task solvability.

Each of the five conditions contains 12 panels, two per family. Fresh Dev retains the sampler's name but is evaluation-only, just like the other four conditions. Checkpoint selection uses the separate demonstration-development split. The original rollout-development scenes are used only for the labeled, development-informed description controls in Appendix E.2. They are not additional fresh scenes.

B.3 FAMILIAR-START COMPETENCE

Before execution, we select the lexicographically first successful demonstration in each supported task-family/exact-body stratum. The eight starts comprise two Reach, one Grasp, one Force, two Portal, and two Coupling instances. Both observation-history slots, validity masks, goals, and body inputs match their training records for all 48 executions. Table 7 gives the complete results. These deterministic positive starts diagnose competence under familiar initial conditions; they are not a sample-based estimate of training-set success.

Table 7: Success on the same eight training-support starts for all six checkpoints, under the standard closed-loop budget.

Method	Reach	Grasp	Force	Portal	Coupling	All
HPT-13	1/2	0/1	0/1	0/2	0/2	1/8
DP-13	2/2	0/1	1/1	1/2	0/2	4/8
HPT-37	1/2	0/1	0/1	0/2	0/2	1/8
DP-37	0/2	1/1	1/1	1/2	0/2	3/8
HPT-71	0/2	0/1	0/1	0/2	0/2	0/8
DP-71	2/2	0/1	0/1	0/2	0/2	2/8

C POLICY TRAINING AND COMPUTATION

Both methods use two observation frames, an eight-command prediction horizon, and at most four executed commands before replanning. Training uses AdamW (Loshchilov & Hutter, 2019), batch size 16, learning rate 10^{-4}, weight decay 10^{-6}, $(\beta_1, \beta_2) = (0.95, 0.999)$, and gradient clipping at norm 1. Both diffusion heads predict epsilon with a squared-cosine schedule and 100 training timesteps; inference uses 100 steps. Their method-specific recipes are fixed before fresh evaluation.

HPT. HPT uses a frozen 16-block transformer with width 256 and eight heads. The frozen ImageNet ResNet-18 receives images resized to 224×224. Domain stems and a conditional U-Net head are trained with constant learning rate. The head has channels $[64, 128]$, diffusion-step embedding width 32, kernel size 3, eight groups, and a 256-dimensional global condition; inference uses the author's DDIM sampler.

Diffusion Policy. Diffusion Policy uses trainable ImageNet-initialized ResNet-18 encoders and 116×116 crops from the 128×128 views. Its conditional U-Net has channels $[512, 1024, 2048]$, diffusion-step embedding width 128, kernel size 5, eight groups, and global observation conditioning. Training uses cosine learning-rate scheduling; inference uses EMA weights and DDPM with fixed-small variance. Both methods start from the same ImageNet ResNet-18 visual weights, but HPT additionally uses its pretrained shared transformer.

HPT fits normalizers once on pooled training data. DP uses fixed public feature units and global action bounds. Output unnormalization is shared across bodies and unchanged at evaluation. Body-description interventions therefore modify policy inputs rather than the action rescaling.

All six runs complete 50 epochs. Demonstration-development diffusion loss selects epochs 49, 43, and 47 for HPT seeds 13, 37, and 71, and 38, 33, and 36 for DP. Selection does not shorten training and uses no rollout outcome. Completing the schedule does not establish optimization convergence.

Each run uses one NVIDIA RTX A6000. Recorded training-and-validation epoch times sum to 76.3 hours: 14.2–15.0 per HPT seed and 10.4–11.4 per DP seed. They exclude setup, checkpoint writing, and interruption downtime. The models have 3,368,552 and 338,546,952 trainable parameters, excluding frozen weights. Across the six development matrices (144 episodes, 14,400 queries), per-seed median query times are 352–356 ms and 717–721 ms. Queries include public-snapshot recording. Simulation has no real-time inference deadline, and shared-host timing is not a controlled speed comparison.

D COMPLETE CHALLENGE-SUITE RESULTS

Tables 8 and 9 retain all assigned members, including failures and families without demonstrations. Each family has four members per seed and condition; one outcome changes a seed's family rate by 25 percentage points. The three seeds share the same scenes.

Table 8: Task successes out of 24 for every fixed checkpoint and fresh condition. Each checkpoint is selected after a complete 50-epoch run.

Method	Seed	Fresh Dev	Public test	Body OOD	Scene OOD	Body+Scene OOD
HPT	13	2/24	2/24	1/24	0/24	1/24
HPT	37	0/24	1/24	0/24	0/24	0/24
HPT	71	0/24	0/24	0/24	0/24	0/24
Diffusion Policy	13	6/24	5/24	5/24	0/24	0/24
Diffusion Policy	37	5/24	6/24	3/24	0/24	1/24
Diffusion Policy	71	4/24	3/24	4/24	0/24	1/24

Table 9: Family success (%, mean $\pm$ sample SD across three seeds). Each condition uses two shared scenes and four members per family. Different conditions draw different scenes.

Method	Condition	Reach	Grasp	Force	Portal	Spool	Coupling
HPT	Fresh Dev	8.3 ± 14.4	8.3 ± 14.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
HPT	Public test	16.7 ± 28.9	8.3 ± 14.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
HPT	Body OOD	0.0 ± 0.0	8.3 ± 14.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
HPT	Scene OOD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
HPT	Body+Scene OOD	8.3 ± 14.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Diffusion Policy	Fresh Dev	83.3 ± 14.4	0.0 ± 0.0	25.0 ± 25.0	8.3 ± 14.4	0.0 ± 0.0	8.3 ± 14.4
Diffusion Policy	Public test	83.3 ± 28.9	33.3 ± 14.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Diffusion Policy	Body OOD	75.0 ± 25.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	25.0 ± 0.0
Diffusion Policy	Scene OOD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Diffusion Policy	Body+Scene OOD	16.7 ± 14.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

D.1 ATTAINABILITY AND TRAINING SUPPORT

The fixed privileged expert attempts every member once with the same 400-command budget. Its recipes use exact geometry and simulator state but execute through native contact physics, without object attachment or post-reset pose changes. It succeeds on 57/120 members, with checker agreement and no execution errors. Of the 63 failures, 41 accompany a planning or budget stop and 22 follow program completion. Two successful Coupling executions satisfy the predicate before a later program stop, illustrating why program completion and task success are separate.

Table 10: Expert successes, out of four members per family and 24 per condition. These privileged executions provide attainability witnesses rather than equal-information baseline scores.

Family	Fresh Dev	Public test	Body OOD	Scene OOD	Body+Scene OOD	All
Reach	4/4	4/4	4/4	4/4	4/4	20/20
Grasp	2/4	2/4	2/4	2/4	2/4	10/20
Force	2/4	2/4	2/4	2/4	1/4	9/20
Portal	2/4	4/4	2/4	0/4	0/4	8/20
Spool extraction	0/4	0/4	0/4	0/4	0/4	0/20
Grasp–portal coupling	4/4	4/4	2/4	0/4	0/4	10/20
Total	14/24	16/24	12/24	8/24	7/24	57/120

Every Spool attempt stops at the far-side pregrasp IK query. At those prescribed targets, the terminal arm link intersects a wall under the planner collision model; its fixed link-to-tool transform makes that pose independent of the IK initialization. Convex-mesh checks against the four wall boxes find penetration above 1 μm for all 20 members, with transforms checked against saved forward kinematics. This explains failure of the fixed pregrasp recipe, not impossibility of another grasp or of the task.

Witness-positive and witness-unresolved members partition the unchanged roster. On the 57 witnessed members, HPT succeeds 6, 1, and 0 times across seeds, while DP succeeds 15, 14, and 11 times. The remaining 63 members account for no HPT success and one DP success per seed on two members of one Body-OOD Coupling panel. These policy successes supply additional evidence of attainability without retrospectively redefining the expert strata.

Table 11: Policy success by expert-witness status. Counts pool three checkpoints per architecture over 120 distinct members; checkpoint repetitions are not independent tasks.

Expert-witness stratum	Members	HPT (3 seeds)	DP (3 seeds)
All members	120	7/360	43/360
Positive witness	57	7/171	40/171
No expert witness	63	0/189	3/189

Training support compares each evaluation body's numeric physical description with TRAIN descriptions for its task, excluding development demonstrations. Of 48 exactly matched members, 38 also have expert witnesses. HPT succeeds in 5/114 and DP in 29/114 executions on these doubly supported members. Scene difficulty still varies within these strata, so the rates do not estimate a causal effect of supervision.

D.2 Joint evidence on complete scene pairs

Let E_k denote panels with exactly k expert-positive members, and T_2 denote panels with exact TRAIN-body support for both members. These retrospective labels use expert outcomes and training metadata, not learned-policy success. Table 12 partitions all 60 scenes, retaining the empty stratum. The 18 E_2 pairs include 13 in T_2: six Reach, three Portal, and four Coupling scenes. The other five remain evaluated and contribute 11/30 DP matching successes. Grasp and Force have no E_2 panel; Spool has neither demonstration support nor an expert witness.

Table 12: Complete expert-witness/training-support partition. Matching counts pool three checkpoints and both bodies. Shared counts denote directional expert sequences, not checkpoint-expanded copies.

Stratum	Scenes	HPT matching	DP matching	Shared expert sequences / sources
All	60	7/360	43/360	36/120
$E_2,\ T_2$	13	3/78	22/78	26/26
$E_2,\ \neg T_2$	5	1/30	11/30	10/10
$E_1,\ T_2$	0	–	–	–
$E_1,\ \neg T_2$	21	3/126	7/126	0/42
$E_0,\ T_2$	5	0/30	0/30	0/10
$E_0,\ \neg T_2$	16	0/96	3/96	0/32

Table 13: Joint outcomes on all 13 $E_2 \cap T_2$ scenes. Every architecture has 39 four-cell blocks and 78 matching-description executions on these same pairs.

Policy	Four-cell successes $(00, 01, 10, 11)$, each /39	Matching pair types both/only0/only1/neither	All-zero blocks	$C < 0/C = 0/C > 0$ blocks	Mean C (pp)
HPT	$(2, 2, 3, 1)$	$(1, 1, 0, 37)$	36/39	$(1, 38, 0)$	-2.56
DP	$(11, 10, 12, 11)$	$(9, 2, 2, 26)$	25/39	$(1, 37, 1)$	0.00

DP's 22 matching successes comprise Reach 20/36, Portal 1/18, and Coupling 1/24. Its mean $C = 0$ combines one positive, one negative, and 37 zero blocks; 25/39 blocks fail in all four cells. HPT succeeds in 3/78 matching executions and has 36/39 all-zero blocks. Table 14 lists every nonzero block, not only the positive example discussed in the main text. Each listed scene has unchanged shared expert sequences, including the negative-benefit cases. Shared solutions and policy-specific matching benefit therefore remain distinct even within a fixed support stratum.

Table 14: All nonzero-C blocks in the 13-scene support stratum. C is the unscaled contrast. Each scene has two shared expert-sequence witnesses; zero and failed blocks remain in Table 13.

Split	Family	Scene index	Checkpoint	$(Y_{00}, Y_{01}, Y_{10}, Y_{11})$	C	Shared expert witnesses
Fresh Dev	Portal	1	DP-71	$(1, 0, 0, 0)$	+0.5	2
Fresh Dev	Coupling	1	DP-37	$(0, 0, 1, 0)$	-0.5	2
Public test	Portal	1	HPT-37	$(0, 1, 1, 0)$	-1.0	2

E Description interventions and recorded-action replay

E.1 Complete body–description crossing

For a fixed scene and checkpoint, Y_{bd} denotes success with physical body b and supplied description d. The matching rollouts provide Y_{00}, Y_{11}; reciprocal descriptions provide Y_{01}, Y_{10}. Physical resets, goals, seeds, checkpoints, normalizers, and initial non-description observations are fixed within a physical member. Subsequent observations may differ as actions diverge. Each execution retains 400 commands, 100 queries, two history positions, and 100 diffusion steps per query.

At fixed physics, $D_b = Y_{b1} - Y_{b0}$ measures description response; at fixed description, $P_d = Y_{1d} - Y_{0d}$ measures physical response with feedback. Their interaction and the original diagonal response satisfy

$$I = Y_{11} - Y_{10} - Y_{01} + Y_{00} = D_1 - D_0 = P_1 - P_0 = 2C,$$

$$Y_{11} - Y_{00} = \frac{P_0 + P_1}{2} + \frac{D_0 + D_1}{2}.$$

Thus I and C are the same contrast at different scales, and the diagonal response does not isolate body metadata. RGB and proprioception remain available; this experiment is not a retraining ablation without descriptions.

The exploratory design is fixed after the original cohort outcomes are available. The factorial evaluation contains 1,440 cells in 360 checkpoint–scene blocks. Checked failures are zero; execution errors or unfinished slots are missing rather than failures. One renderer-construction error occurred before any query or command. We reran that same slot without changing the scientific settings; the rerun failed and contributes zero, while the original error remains in the released record. No scene was replaced based on an outcome.

Empirical intervals use the central 95% of 10,000 bootstrap estimates with seed 20260915. Whole scenes are resampled within split–family strata, retaining all four cells and all six checkpoints. These intervals describe resampling variation conditional on the observed scenes and checkpoints. With only two scenes per stratum, they do not establish nominal 95% coverage for a new-scene population; degenerate intervals are withheld and fine-grained results are descriptive.

Matching and reciprocal descriptions yield 7/360 versus 9/360 HPT successes, and 43/360 versus 46/360 DP successes. Tables 15 and 16 give the cell means and per-seed effects. The near-zero mean matching benefits coexist with nonzero physical contrasts.

E.2 Development-informed feature controls

Seed-13 HPT and DP are also evaluated on all 24 original development members under identity, paired-description swap, global shuffle, and feature-zero controls. These 192 control executions use the same physical resets and are separate from 48 earlier baseline executions. Settings are fixed after

Table 15: Reciprocal-description factorial on 60 paired scenes. Estimates pool three selected checkpoints per architecture (training seeds 13, 37, and 71). Y_{bd} indexes physical member b and supplied description d, retaining the original member orientation.

	HPT		Diffusion Policy	
Quantity	Estimate	Emp. 95% interval	Estimate	Emp. 95% interval
Success (%)				
Y_{00}	1.7	[0.6, 2.8]	9.4	[7.2, 11.7]
Y_{01}	1.1	[0.0, 2.2]	10.0	[7.8, 12.2]
Y_{10}	3.9	[1.7, 6.1]	15.6	[12.8, 18.9]
Y_{11}	2.2	[0.6, 3.9]	14.4	[12.2, 16.7]
Effects (percentage points)				
$D_0 = Y_{01} - Y_{00}$	-0.6	[−1.7, 0.6]	0.6	[−1.7, 2.8]
$D_1 = Y_{11} - Y_{10}$	-1.7	[−2.8, −0.6]	-1.1	[−3.9, 1.7]
$P_0 = Y_{10} - Y_{00}$	2.2	[0.6, 3.9]	6.1	[3.9, 8.3]
$P_1 = Y_{11} - Y_{01}$	1.1	[−0.6, 2.8]	4.4	[3.3, 5.6]
$C = I/2$ (matching)	-0.6	[−1.1, 0.0]	-0.8	[−1.9, 0.3]
$Y_{11} - Y_{00}$ (diagonal)	0.6	[−1.1, 2.2]	5.0	[2.8, 7.2]

Each cell mean uses 180 binary outcomes per architecture; each contrast uses 180 matched four-cell blocks on the same 60 scenes. All 1,440 runs are checked: 720 original diagonals and 720 reciprocal-description swaps. Intervals span the saved central 95% of paired-scene bootstrap estimates (10,000 draws, seed 20260915), stratified by split × family with all four cells and all six checkpoints retained per scene. They describe resampling variation conditional on the empirical scenes and selected checkpoints. Two scenes per stratum do not establish nominal 95% new-scene coverage. Matching benefit C is the primary descriptive summary; $I = 2C$ is the same contrast at twice the scale.

those baseline outcomes are known. The global shuffle uses one permutation with seed 20260918; all donors are retained, including six recipients whose encoded features remain identical. Feature zero replaces the 24 raw body features before the checkpoint's fixed normalizer. It removes only that explicit body-feature input, not body information in vision or proprioception.

Initial non-description inputs match across conditions. We compare the first acknowledged action at that shared state; later action changes can also reflect feedback. Identity reproduces all 400 baseline commands for every member. The paired swap changes every first action, while global shuffle changes 18/24 and feature zero 24/24 for each method. Table 17 reports both numerical action sensitivity and outcomes.

Paired swapping changes HPT success from 2/24 to 3/24 and DP from 4/24 to 5/24. HPT has three gains and two losses; DP has one gain and no loss. HPT's zero-feature total remains 2/24 while hiding one gain and one loss; DP retains the same four successful members under shuffle and zero despite changed initial commands. Action dependence, aggregate success, and member-level success therefore need not agree. These are fixed-checkpoint controls, not separately trained metadata-hidden policies.

E.3 Fixed-command replay

All 120 fresh-cohort expert sequences, including 63 failed attempts, are replayed on their original and paired bodies. Each retains its original 102–400 commands, reset, goal, seed, and control rate, with no padding or early stopping. Source expert outcomes are known before replay; the control is a posthoc diagnostic, not an extra policy baseline.

Every source satisfies the shared global command bounds. Body-range compatibility is checked separately. The 110 compatible sequences retain native float32 commands. The other ten are projected to destination joint/jaw ranges, with original and executed targets both retained. Their unprojected transfers are not measured. Identity reproduces all source outcomes and complete joint/object trajectories, and both checkers recompute all 240 replay outcomes.

Transfer succeeds on 45 unchanged sequences and no projected sequence. Across all sources, 36 succeed on both bodies, 21 only on the source, nine only after transfer, and 54 on neither. All 36 directional sequences from the 18 two-sided expert-positive panels transfer unchanged and success-

Table 16: Training-seed detail for the reciprocal-description factorial. Each seed uses the same 60 paired scenes (60 outcomes per Y cell and 60 four-cell blocks per contrast). Y cells are percent success; contrasts and all SDs are in percentage points.

Architecture	Quantity	Seed 13	Seed 37	Seed 71	Mean	Seed SD
HPT	Y_{00}	5.0	0.0	0.0	1.7	2.9
	Y_{01}	1.7	1.7	0.0	1.1	1.0
	Y_{10}	8.3	3.3	0.0	3.9	4.2
	Y_{11}	5.0	1.7	0.0	2.2	2.5
	D_0	-3.3	1.7	0.0	-0.6	2.5
	D_1	-3.3	-1.7	0.0	-1.7	1.7
	P_0	3.3	3.3	0.0	2.2	1.9
	P_1	3.3	0.0	0.0	1.1	1.9
	C	0.0	-1.7	0.0	-0.6	1.0
	$Y_{11} - Y_{00}$	0.0	1.7	0.0	0.6	1.0
Diffusion Policy	Y_{00}	10.0	10.0	8.3	9.4	1.0
	Y_{01}	11.7	8.3	10.0	10.0	1.7
	Y_{10}	16.7	16.7	13.3	15.6	1.9
	Y_{11}	16.7	15.0	11.7	14.4	2.5
	D_0	1.7	-1.7	1.7	0.6	1.9
	D_1	0.0	-1.7	-1.7	-1.1	1.0
	P_0	6.7	6.7	5.0	6.1	1.0
	P_1	5.0	6.7	1.7	4.4	2.5
	C	-0.8	0.0	-1.7	-0.8	0.8
	$Y_{11} - Y_{00}$	6.7	5.0	3.3	5.0	1.7

Mean and sample SD are the saved summaries of the three per-seed means. Seed SD describes training-seed variation on these fixed scenes; it is not a confidence interval and the seeds do not increase the number of independent scenes. D_0, D_1 change the supplied description at fixed physics; P_0, P_1 change the physical body at fixed description, allowing closed-loop policy responses. $C = I/2$; the original diagonal response is $Y_{11} - Y_{00}$.

Table 17: Body-feature controls on the original development roster. Success is out of 24 members. First-action changes use tolerances 10^{-6} rad or 10^{-6} m and measure numerical sensitivity. Joint differences average seven joints and 24 members; jaw differences are separate.

Method	Metadata	Changed	Joint Δ (rad)	Jaw Δ (mm)	Success
HPT-13	Correct	0/24	0.000000	0.000000	2/24
HPT-13	Shuffled	18/24	0.003553	0.099851	1/24
HPT-13	Incorrect (paired)	24/24	0.000848	0.045672	3/24
HPT-13	Zero features	24/24	0.050341	4.540407	2/24
DP-13	Correct	0/24	0.000000	0.000000	4/24
DP-13	Shuffled	18/24	0.004172	0.054205	4/24
DP-13	Incorrect (paired)	24/24	0.000627	0.012387	5/24
DP-13	Zero features	24/24	0.009093	0.242957	4/24

fully. If $\mathcal{S}_b$ is the set of successful global-command-valid sequences for a body and reset, each shared sequence witnesses $\mathcal{S}_0 \cap \mathcal{S}_1 \neq \varnothing$. Failed transfers cannot establish an empty intersection, and a witness does not show that a learned policy can find it.

F Calibration with public-input controllers

The calibration cohorts are separate from the learning demonstrations and the 60 challenge scenes. Both use analytic controllers operating on public RGB-D and measured robot state. They test the meaning of the diagnostics, not architectural superiority over the RGB-only learned baselines.

Table 18: Replay of all 120 sources, retaining source failures. Shared sequences are unchanged and succeed on both bodies. Ten projected transfers have no measured unprojected counterpart.

Source-sequence subset	N	Identity success	Transfer success	Shared sequence
All source sequences	120	57/120	45/120	36/120
Unchanged on paired body	110	47/110	45/110	36/110
Body-projected on paired body	10	10/10	0/10	0/10
Both-source-positive (18 panels)	36	36/36	36/36	36/36

F.1 Original Portal: competent execution with shared solutions

The whole-path reference retains the original nominal Portal task, wrist-guard pair, carried-envelope passage, stable released placement, and contact restrictions. Controller A uses the supplied complete body description. Controller S substitutes the nominal b_1 description before reading or validating the donor. Runtime inputs are two calibrated RGB-D views, measured joint/jaw histories, the public goal, and body description. Offline development additionally uses success, passage, forbidden-collision, and grasp-loss summaries; these are unavailable at runtime.

Perception and control. The observer estimates object and marker bounds, support plane, and wall slab. Observed jambs, sill edges, and rays reaching behind the slab support a rectangular-prismatic completion of partly unobserved aperture interiors. Lintel observations can extend the supported opening if through-rays agree and observed wall points do not contradict it. Missing lateral support prevents route construction; slab intersections outside supported openings are occupied. This completion is a scene prior, not a free-space certificate.

After pregrasp and closure, the planner expresses the observed object relative to the measured tool frame and plans transport before lifting. Waypoints cover lift, alignment, six crossing stations, clearance, placement, and withdrawal. Release postures are checked first. A layered roadmap retains at most seven IK solutions per waypoint and a four-path beam, testing direct joint edges before a fixed-seed RRT-Connect fallback (Kuffner & LaValle, 2000). Separate pregrasp and transport phases each allow 400 IK solves and 20,000 collision checks, with at most 500 iterations per fallback connection. The unchanged 400-command horizon reserves opening, settling, and feedback.

Checks transform the supplied convex body shapes, finger envelopes, and carried-object proxy. Wall tests clip convex sections to the slab; table checks depend on motion phase, and self-collision checks cover selected nonadjacent arm groups. Joint edges are sampled at increments no greater than 0.035 rad, with a 12 mm wall margin and 3 mm object inflation. Feedback limits increments to 0.12 rad per query, checks a 3 mm tracking margin, and allows bounded reconnection (100 iterations per connection, 2,000 collision checks per episode). Planning failure produces a measured-arm hold with a closed jaw while carrying. The sampled geometry and assumed rigid object–tool transform do not guarantee continuous collision freedom or grasp retention.

Development and reserved evaluation. Fourteen development pairs are exposed during reference development. The eight qualification pairs and sixteen evaluation pairs are reserved beforehand and provide no tuning observations. A four-scene functional gate fixes the final reference; it requires at least 3/4 matching successes per controller–body stratum, bilateral success on at least two scenes per controller, valid slots, donor-invariant S, and same-body replay identity. On the reserved eight pairs, the reference must admit every pair, provide witnesses for both bodies, complete every prescribed execution, and reach at least 7/8 successes in each matching controller–body condition. A body's witness is the first checked success in the fixed order: privileged expert, matching A, then matching S. E_2 denotes expert success on both bodies; W_2 includes successful references under this rule.

Every reserved evaluation pair receives all eight controller–description cells regardless of expert success. The full-range criterion requires all sixteen pairs admitted and witnessed, all 128 controller cells valid, descriptor-invariant S, and at least 14/16 matching successes per condition. This criterion describes the reserved range; it does not remove rows from the reported evaluation. Each pair also supplies a fixed S_{00} identity replay and an unchanged paired-body transfer. Positive matching benefit is not required.

Quantity	Qualification ($N = 8$)	Evaluation ($N = 16$)
Admitted pairs	$8/8$	$16/16$
A_{00} successes	$8/8$	$16/16$
A_{11} successes	$8/8$	$16/16$
S_{00} successes	$8/8$	$16/16$
S_{11} successes	$8/8$	$16/16$
W_2 pairs	$8/8$	$16/16$
E_2 pairs	$3/8$	$12/16$
Observed controller cells	$32/32$	$128/128$
Complete prescribed slots	$64/64$	$224/224$
Bound, agreeing checked action traces	$48/48$	$192/192$
Phase outcome	complete	complete
Mean C_A	–	0
Mean C_S	–	0

Table 19: Original Portal qualification and evaluation. All prescribed outcomes are complete. W_2 uses checked successes from the fixed expert and matching-reference sources; E_2 requires expert success on both bodies. Verification repeats are separate from these rosters.

The reference meets the reserved qualification criteria. In evaluation, every controller cell succeeds, $C_A = C_S = 0$ on every pair, and all sixteen sequences succeed in both replays. Of the 32 physical members, the witness rule selects 27 expert and five matching-A successes; expert-only panel counts are $E_0 = 1, E_1 = 3, E_2 = 12$. The references therefore extend observed attainability beyond the fixed expert. Table 19 reports the development and evaluation cohorts separately.

Verification and contact-order check. A saved-source audit accepts 143/144 sources before stopping on a qualification expert failure whose simultaneous contacts appear in a different order on recomputation. Both unchanged checkers preserve the failure and all 533 collision events. We allow contact permutations only within one physics timestep when event values and multiplicities agree; cross-timestep order and all other fields remain unchanged.

Verification then uses the four evaluation pairs selected in advance, without policy changes or replacement. All 48 repeated action episodes succeed; the 40 controller/replay comparisons reproduce initial inputs, commands, and outcomes, and four identity replays reproduce native traces exactly. These repeats add no new evaluation scenes. The result establishes competent execution on the nominal Portal sampler, not on every family or on hardware.

Earlier Portal controller variants. Two earlier public-input references using the same nominal Portal sampler did not meet the prespecified development criteria. One completed 48 development executions with checker agreement but scored $0/12$ in every matching stratum. A second completed its functional batches but missed the 3/4 matching threshold and, for one constrained variant, the bilateral S threshold. These variants were not evaluated on the reserved qualification or test scenes; they are development records, not controlled ablations.

F.2 Constructed Grasp: matching benefit with shared solutions

The two bodies have 35/80 mm maximum openings, identical arm limits, 100 N per-finger caps, and no guard. A $22 \times 60 \times 40$ mm box has mass 0.15 kg and friction 1.0. The reset is $(0, \pi/8, 0, -5\pi/8, 0, 3\pi/4, \pi/4)$ with each finger at 0.01 m. Two development positions are $(-0.14, -0.14)$ and $(-0.10, -0.10)$ m. Evaluation uses the twelve-point grid $x \in \{-0.16, -0.12, -0.08\}$, $y \in \{-0.16, -0.12, -0.08, -0.04\}$ m. Geometry, order, seeds, controllers, and criteria are fixed before execution. All 28 physical-member resets pass admission.

A color/shape observer estimates object bounds from the public RGB-D and calibration. A tests Y-closing then X-closing grasps, selecting the first whose observed width plus 6 mm fits the supplied opening; its IK model also uses the supplied anatomy. S always takes the short-side grasp and constant 35 mm anatomy. Both use measured-joint feedback, with no private object state or reward. Six stages allocate 80, 20, 40, 80, 40, and 40 commands; a final hold fills the 400-command budget.

Table 20: Constructed 12-scene Grasp calibration with fixed RGB-D analytic controls. Physical body b and supplied description d are crossed in every scene. Each cell entry is checked successes out of 12; $100\bar{C}$ is the matching-description contrast in percentage points. Results describe this fixed cohort, without population confidence intervals.

Control	Y_{00}	Y_{01}	Y_{10}	Y_{11}	$100\bar{C}$ (pp)
A	10/12	0/12	10/12	12/12	50.0
S	10/12	10/12	10/12	10/12	0.0

A uses the supplied anatomy to select a grasp and construct its IK commands; S fixes the short-side grasp and the public 35 mm anatomy. Unchanged S00 replay succeeds in 10/12 identity and 10/12 transfer executions; shared-sequence witnesses occur in 10/12 scenes, including 10/12 with positive matching benefit for A. Successful-witness coverage is 22/24 body-scene members. The prespecified joint calibration criterion is not met. Members `calib_eval_07_b0`, `calib_eval_11_b0` have no successful prespecified witness.

IK uses orientation/posture weights 0.1/0.0001, maximum increment 0.08 rad, at most 48 function evaluations, and tolerance 10^{-7}.

Development requires all four matching-A successes with positive C_A in both scenes, all eight S successes, and donor-invariant S commands. These conditions pass. Evaluation contains 96 controller cells and 24 unchanged S_{00} replays. The full-range criterion requires at least 10/12 matching successes per body for both controllers, mean $C_A \geq 0.25$, at least six positive and no negative scene contrasts, $C_S = 0$ with command invariance, witnesses for all 24 physical members, and a shared replay with positive C_A.

All 120 evaluation executions complete: 92 succeed and 28 fail. Every scene has $C_A = 0.5$ and $C_S = 0$; ten have shared successful commands. The two narrow-body members at $(-0.12, -0.04)$ and $(-0.08, -0.04)$ m lack success witnesses, leaving 22/24 witnessed members. The full-range criterion is therefore not met, while all rows remain in the reported denominator.

F.3 A COMPLETED LEARNED-REFERENCE ATTEMPT

A separate recurrent imitation reference tests whether successful Portal supervision yields competent learned execution. Public RGB-D preprocessing supplies geometry estimates and missingness masks, concatenated with proprioception, goal, and body features. Two width-128 ReLU layers feed a width-128 GRU (Cho et al., 2014) and an eight-output head. Arm increments are limited to 0.12 rad; each query repeats its target for four commands. Runtime uses the network without teacher planning or fallback.

On 32 separate training pairs, an analytic teacher generates 96 successful 400-command demonstrations (A_{00}, A_{11}, S_{00}), giving 9,600 query features. One seed-13 student trains for 50 full-sequence epochs with AdamW, learning rate 3×10^{-4}, weight decay 10^{-4}, batch size eight episodes, and gradient clipping at one. Squared arm/jaw errors are scaled by 0.12 rad and 0.12 m. The final epoch is retained without development-based selection.

All 24 executions on four fixed development pairs fail, including $0/4$ in each matching condition. The required threshold is at least 3/4 per matching condition and two bilateral successful pairs per controller, so no reserved evaluation is run. This result bounds the calibration: successful teacher trajectories and completed optimization did not yield learned closed-loop competence.

G REPRODUCIBILITY AND RELATION TO POLICY ALTERNATIVES

The supplementary material contains corrected demonstrations, task definitions, training/evaluation procedures, and recorded executions. Pretrained-component redistribution remains unresolved; the six selected checkpoints are inspectable. The calibration cohorts, challenge evaluation, and training demonstrations are distinct resources.

Completed reproducibility checks. Loading recovers all 100 episodes, both window counts, and 120 evaluation requests. On one real training window per checkpoint, matched-noise CPU inference reproduces inputs, normalizers, weights, and predictions bit-for-bit. Two closed-loop repetitions,

HPT seed 13 at epoch 49 and DP seed 13 at epoch 38, use the first fresh-development member fixed before reproduction. Each reproduces 400 commands, 100 queries, observations, full traces, and checked outcomes. These checks cover execution and aggregate reconstruction; they are not fresh installation or full retraining.

Reported aggregates are also reconstructed from saved execution evidence, including the 720 matching rollouts, 48 familiar-start executions, 192 feature controls and 48 baselines, and 240 expert replays. An example pair contains both public observation streams, 400 commands, and 2,001 physics samples per rollout. Recorded-aggregate reconstruction and newly executed command replay are distinct checks.

Observation and learning alternatives. PerAct and DP3 use depth-derived voxel and point-cloud representations (Shridhar et al., 2023; Ze et al., 2024); recorded depth does not make an RGB-only baseline equivalent to these approaches. RMA explicitly learns online adaptation (Kumar et al., 2021). Visual and dynamics randomization target different distribution shifts (Tobin et al., 2017; Peng et al., 2018). BodyBound supplies controlled comparisons for such future studies, while the present evidence concerns its stated observations, supervision, and fixed simulation budgets.